

Michigan Law Review

Volume 89 | Issue 6

1991

Statistics for Lawyers and Law for Statistics

D. H. Kaye

Arizona State University College of Law

Follow this and additional works at: <https://repository.law.umich.edu/mlr>



Part of the [Legal Profession Commons](#)

Recommended Citation

D. H. Kaye, *Statistics for Lawyers and Law for Statistics*, 89 MICH. L. REV. 1520 (1991).

Available at: <https://repository.law.umich.edu/mlr/vol89/iss6/19>

This Review is brought to you for free and open access by the Michigan Law Review at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Michigan Law Review by an authorized editor of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

STATISTICS FOR LAWYERS AND LAW FOR STATISTICS

*D.H. Kaye**

STATISTICS FOR LAWYERS. By *Michael O. Finkelstein* and *Bruce Levin*. New York: Springer-Verlag. 1990. Pp. xxii, 608. Cloth \$59; paper \$39.

Statistics is an emotionally charged subject. Many people fear it and hate it; some revel in it and worship it. Inside the discipline the Montagues engage in wordy debates with the Capulets, and sometimes even sue them in courts for holding wrong opinions. Most statisticians despise other statisticians, and all despise all nonstatisticians, who would like to be able to return the compliment, but are not quite sure how to shrug off a "best fit." In an atmosphere of mutual suspicion and contempt, neither side is willing to learn what the other has to teach.¹

These observations, penned some twenty years ago, retain their truth today, and they apply with special force to one prominent subset of nonstatisticians — lawyers. Nevertheless, even within a profession as nonquantitatively inclined as law, exceptions exist. Michael Finkelstein is one. An accomplished lawyer and adjunct professor at Columbia Law School, Finkelstein has spent at least a quarter century learning what statisticians have to teach and applying this knowledge in litigation and legal education.² He knows quite well how to shrug off, or put on, a best fit. In *Statistics for Lawyers*, Finkelstein teams up with Bruce Levin, a biostatistician at Columbia University's School of Public Health, to produce a unique textbook on statistics — a mixture

* Regents Professor, Arizona State University College of Law. S.B. 1968, MIT; A.M. 1969, Harvard; J.D. 1972, Yale. — Ed. I am grateful to William Fairley and Hans Zeisel for comments on a draft of this paper and to Tony D'Amato for correspondence and copies of portions of the transcript and the appellant's brief in *Branion v. Gramly*, 855 F.2d 1256 (7th Cir. 1988). Portions of this review were completed while the author was a Visiting Research Fellow of the University of Chicago School of Law.

1. J. LUCAS, *THE CONCEPT OF PROBABILITY* 163 (1970).

2. See M. FINKELSTEIN, *QUANTITATIVE METHODS IN LAW: STUDIES IN THE APPLICATION OF MATHEMATICAL PROBABILITY AND STATISTICS TO LEGAL PROBLEMS* (1978); Finkelstein, *The Application of Statistical Decision Theory to Jury Discrimination Cases*, 80 HARV. L. REV. 338 (1966); Finkelstein, *The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases*, 80 COLUM. L. REV. 737 (1980); Finkelstein, *Multiple Regression Models in Employment Discrimination Cases — The Problem of Imperfect Proxies*, 31 JURIMETRICS J. 109 (1990); Finkelstein & Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970); Finkelstein & Levenbach, *Regression Estimates of Damages in Price-Fixing Cases*, LAW & CONTEMP. PROBS., Autumn 1983, at 145; Finkelstein & Robbins, *A Probabilistic Approach to Tracing Presumptions in the Law of Restitution*, 24 JURIMETRICS J. 65 (1983).

of legal vignettes, doctrine, statistical concepts, and exercises in applying these concepts.

In this review, I describe and assess this unusual enchiridion. Part I summarizes the book's objectives and accomplishments. I conclude that any prospective or practicing attorney who works through *Statistics for Lawyers* will indeed emerge with "a fuller appreciation of the standards for analyzing data and making inferences" (p. ix). Part II applies some of these standards to the use of probability theory in *Brannon v. Gramly*.³ I criticize the mathematical argument advanced in that case and suggest that its defects would have been apparent to lawyers with the "fuller appreciation" that *Statistics for Lawyers* seeks to convey.

I. STATISTICS FOR LAWYERS

Cases of false advertising,⁴ theft,⁵ mass murder,⁶ discrimination by race, age, and gender,⁷ voting rights,⁸ antitrust,⁹ taxation,¹⁰ domestic relations,¹¹ environmental torts,¹² and product liability¹³ dance across

3. 855 F.2d 1256 (7th Cir. 1988).

4. Pp. 34-37 (drug efficacy claims); see also Barnes, *The Significance of Quantitative Evidence in Federal Trade Commission Deceptive Advertising Cases*, LAW & CONTEMP. PROBS., Autumn 1983, at 25.

5. Pp. 37-41 (estimating losses from defalcation); pp. 84-90 (robbery).

6. P. 90; see also Fienberg & Kaye, *Legal and Statistical Aspects of Some Mysterious Clusters*, 154 J. ROYAL STATISTICAL SOC. 61 (1991).

7. Pp. 1-6, 311-12 (employment tests); pp. 59-61, 383-86 (public school finance); pp. 114-17, 207-11 (jury selection); pp. 122-23 (hiring of public school teachers); pp. 126-27 (appointments to city board); pp. 354-56 (promotions of employees); pp. 368-74 (salaries); pp. 404-07 (salaries, hiring and promotions); pp. 408-18 (salaries and promotions); pp. 189-91, 452-67 (capital sentences); see also STATISTICAL METHODS IN DISCRIMINATION LITIGATION (D. Kaye & M. Aickin eds. 1986).

8. Pp. 47-49 (reapportionment); pp. 66-70 (weighted voting); see also Moncrief & Joula, *When the Courts Don't Compute: Mathematics and Floterial Districts in Legislative Reapportionment Cases*, 4 J.L. & POL. 737 (1988).

9. Pp. 70-72 (market concentration guidelines); pp. 386-88 (estimating damages from price-fixing); see also Rubinfeld & Steiner, *Quantitative Methods in Antitrust Litigation*, LAW & CONTEMP. PROBS., Autumn 1983, at 69.

10. Pp. 31-34 (railroad property assessments); pp. 288-89 (valuations of charitable remainders); pp. 290-91 (lifetime of depreciable contracts); see also Bright, Kadane & Nagin, *Statistical Sampling in Tax Audits*, 13 LAW & SOC. INQUIRY 305 (1988); Brown, *Statistical Sampling in the IRS Examination of Large Cases*, TAX EXEC., Apr. 1982, at 175.

11. Pp. 96-101; see also BIOMATHEMATICAL EVIDENCE OF PATERNITY (K. Hummel & J. Gerchow eds. 1982); INCLUSION PROBABILITIES IN PARENTAGE TESTING (R. Walker ed. 1983); Berry, *Inferences Using DNA Profiling in Forensic Identification and Paternity Cases*, STATISTICAL SCI. (forthcoming 1991); Evett & Werrett, *Paternity Calculations from DNA Multilocus Profiles*, 29 J. FORENSIC SCI. SOC. 249 (1989); Kaye, *DNA Paternity Probabilities*, 24 FAM. L.Q. 279 (1990); Kaye, *Presumptions, Probability and Paternity*, 30 JURIMETRICS J. 323 (1990); Kaye, *The Probability of an Ultimate Issue: The Strange Cases of Paternity Testing*, 75 IOWA L. REV. 75 (1989).

12. Pp. 13-22 (leukemia from atomic tests); pp. 66, 297-300 (leukemia from drinking water).

13. Pp. 8-12 (Dalkon shield); pp. 191-93, 255-57 (Bendectin).

the pages of *Statistics for Lawyers*. Many, like *McCleskey v. Kemp*¹⁴ — the powerful but unsuccessful challenge to death sentencing in Georgia — and *People v. Collins*¹⁵ — the misuse of statistics to convict an interracial couple with a yellow car in Los Angeles — already are well-known to many lawyers.¹⁶ Others, like *Anderson v. Cryovac, Inc.*¹⁷ — the out-of-court settlement of the toxic tort claim for contamination of drinking water in Woburn, Massachusetts — have been discussed primarily in the statistical literature.¹⁸ Still others, such as the survey used to determine the sales tax owed by a Colorado mail order firm,¹⁹ have not been reported elsewhere.

As is appropriate to an introductory textbook, the treatment of these cases and issues is panoramic rather than microscopic, and the material is organized around statistical concepts and methods rather than legal subject areas. Nearly every statistical procedure that has proved useful in litigation is described, at least briefly, including many that are not considered in most elementary statistics texts. In addition to the basics of probability theory (pp. 76-110), random sampling (pp. 258-83), descriptive statistics (pp. 1-61), regression (pp. 323-467), and contingency tables (pp. 156-257), *Statistics for Lawyers* describes survival analysis (pp. 284-308), nonparametric methods (pp. 309-22), and jackknife and bootstrap techniques (pp. 461-67). It discusses the hypergeometric distribution (pp. 133-38), which has confused some courts,²⁰ procedures for combining results from related contingency tables (pp. 235-53), which have been overlooked by some courts,²¹ logistic regression (pp. 447-61), which figured prominently in the proof of racial discrimination in *McCleskey v. Kemp*,²² and many biostatistical and epidemiologic concepts, which have become decisively important in ascertaining causation in much litigation and environmental regulation.²³

14. 481 U.S. 279 (1987). Pp. 460.

15. 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968). Pp. 84-88.

16. For newly discovered information about the origin of the probability argument in *Collins*, as well as a fresh treatment of the evidence to which that argument was directed, see Edwards, *Influence, Diagrams, Bayesian Imperialism and the Collins Case: An Appeal to Reason*, CARDOZO L. REV. (forthcoming 1991).

17. 96 F.R.D. 431 (D. Mass. 1982); pp. 66, 297-300; see Butterfield, *Settlement Averts Key Trial in Deaths Tied to Pollution*, N.Y. Times, Sept. 23, 1986, at A16, col. 1.

18. Fienberg & Kaye, *supra* note 6; Lagakos, Wesson & Zelen, *An Analysis of Contaminated Well Water and Health Effects in Woburn, Massachusetts*, 81 J. AM. STATISTICAL A. 583 (1986).

19. Pp. 269-70 (using the case to illustrate nonresponse bias in sampling).

20. Kaye, *Hypergeometric Confusion in the Fourth Circuit*, 26 JURIMETRICS J. 215 (1986).

21. Kaye, *Statistical Evidence of Discrimination in Jury Selection*, in STATISTICAL METHODS IN DISCRIMINATION LITIGATION, *supra* note 7, at 13.

22. 481 U.S. 279 (1987).

23. Compare *Lynch v. Merrill-National Laboratories*, 830 F.2d 1190, 1194, 1196 (1st Cir. 1987) (expert's conclusion that the drug Bendectin is teratogenic was "foundationless" without "confirmatory epidemiologic data") with *DeLuca v. Merrill Dow Pharmaceuticals*, 911 F.2d 941 (3d Cir. 1990) (same expert's opinion was improperly excluded when he presented the same

The statistical expositions that bracket the legal applications shy away from derivations or proofs, and the exercises emphasize conceptualization over computation. Nevertheless, many of the problems are not easily solved, and occasional summation or integral signs in the text may frighten the mathematically faint of heart.²⁴ This would be unfortunate, because the notation and prose are precise, the underlying ideas and their historical foundations are revealed, and the authors' comments are insightful and sophisticated.²⁵

In this respect, the book is an overdue addition to a typically uninspired and oversimplified literature²⁶ in which lawyers (or other non-statisticians trained in economics, social science, or business) try to explain statistics to other lawyers, all too often misstating basic concepts.²⁷ In contrast, *Statistics for Lawyers* is dependably knowledgeable and careful in explaining statistical reasoning and terminology.²⁸

Consider, for example, the book's treatment of screening and diagnostic tests as it applies to a 1987 bill on drug testing in the workplace introduced in the New York Senate. This bill provided that any

epidemiologic findings in the form of confidence intervals). See generally THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS (S. Fienberg ed. 1988) [hereinafter STATISTICAL ASSESSMENTS].

Statistics for Lawyers also discusses meta-analysis, a semiformal technique for combining data from separate studies (pp. 254-57). This procedure, which can be expected to play a controversial role in toxic tort cases, made its forensic debut in *In re Paoli R.R. Yard PCB Litigation*, 916 F.2d 829, 841, 856-58 (3d Cir. 1990).

Of course, not every topic that one might wish to see is included. The ingenious displays of modern exploratory data analysis are not mentioned, and the application of Bayesian inference to problems with continuous variables is slighted.

24. Some paragraphs may elude readers who somehow managed to acquire a college education without calculus. For example, Finkelstein and Levin write that

The cumulative hazard function $H(t)$ at time t is the integral of the hazard function Θ from 0 to t . In continuous time, the hazard function is equal to minus the natural logarithm of the survival function; taking antilogs, the survival function is equal to e raised to the power of minus the cumulative hazard function. In symbols, $H(t) = \int \Theta(u)du = -\log S(t)$, and $S(t) = \exp[-H(t)]$. Thus, one way to estimate $H(t)$ is by summing the discrete hazard estimates d_j/n_j up to time t

P. 286.

25. A variety of statistical issues have puzzled courts. By way of text or problems (happily, with answers collected in an appendix), Finkelstein and Levin address most of them. E.g., pp. 125-26 (arguing that one-tailed tests are more appropriate in most litigation than the more commonly used two-tailed ones); pp. 170-71, 492-93 (criticizing the court's statistical reasoning in *EEOC v. Federal Reserve Bank of Richmond*, 698 F.2d 633 (4th Cir. 1983), *revd. on other grounds*, *Cooper v. Federal Reserve Bank of Richmond*, 467 U.S. 867 (1984)); pp. 212-14, 503-04 (criticizing *Inmates of the Nebraska Penal and Correctional Complex v. Greenholtz*, 567 F.2d 1368 (8th Cir. 1977)); pp. 233-35, 509-10 (criticizing *Moultrie v. Martin*, 690 F.2d 1078 (4th Cir. 1982)).

26. See, e.g., Fienberg, Book Review, 8 J.L. & COMMERCE 317, 321 (1988); Peterson, Book Review, 26 JURIMETRICS 329 (1986).

27. Concepts like P-values and confidence coefficients, for example, are not always defined correctly. Kaye, Book Review, 84 J. AM. STATISTICAL A. 1094 (1989). The descriptions of these terms in *Statistics for Lawyers* are found at p. 124 and pp. 171-81.

28. Another textbook on legal statistics that has this quality is J. GASTWIRTH, STATISTICAL REASONING IN LAW AND PUBLIC POLICY (1988), reviewed by Kaye, *supra* note 27.

screening test "must have a degree of accuracy of at least ninety-five percent" and that "positive test results must then be confirmed by an independent test, using a fundamentally different method and having a degree of accuracy of 98%" (p. 105). This proposed legislation may seem straightforward, but it is not. In general, it takes more than one number to characterize the accuracy of a test: "The *sensitivity* of a test is the proportion of all affected individuals who (correctly) test positive ($P[+|A]$). The *specificity* of a test is the proportion of all unaffected individuals who (correctly) test negative ($P[-|U]$)" (p. 101). There is no necessary connection between these quantities. A test can be exquisitely sensitive (it correctly identifies virtually all the drug users who submit to testing) but not very specific (many nonusers are included among those who test positive); or it can be highly specific but not at all sensitive; or, it can be both extremely sensitive and specific.

The New York bill would have been clearer had it defined the required degree of "accuracy" in terms of sensitivity and specificity. Furthermore, the perceptive reader of *Statistics for Lawyers* will recognize another wrinkle in this seemingly simple bill. As Finkelstein and Levin ask, "Assuming that 0.1% of the adult working population takes drugs, and that 'accuracy' refers to both sensitivity and specificity, what is the positive predictive value of a test program meeting the bill's requirements" (pp. 105-06)? The answer can be found from the formula " $PPV = P[A|+] = P[+,A]/P[+]$," for the "positive predictive value," which is defined as "the proportion of all test-positive-people who are truly affected" (p. 102). Like Finkelstein and Levin, I leave the detailed calculation to the reader, who may be surprised to find that only about one in every fifty-four positives on the screening test with the "95% accuracy," and about one in two positives on both that test and the independent confirmatory test with the "98% accuracy," will be users.²⁹

In raising and analyzing such issues, *Statistics for Lawyers* succeeds admirably in its goal of introducing the ideas and techniques of statistics that have the most application to the courtroom and to the formulation of legal doctrine. Despite its pedestrian title, it is not a routine statistics text with legal examples tossed in. The selection of topics and examples, as well as the exposition of statistics and law, is erudite, informed, and even entertaining. With a strong and steady hand, *Statistics for Lawyers* opens the tool chest of the professional statistician, permitting students of the law to peer within.

29. The small number of true positives results from the low prevalence of drug use in the tested population. This effect of a low base rate is important in many other contexts. See, e.g., W. CURRAN, M. HALL & D. KAYE, *HEALTH CARE LAW, FORENSIC SCIENCE AND PUBLIC POLICY* 163-67 (4th ed. 1990).

II. LAW FOR STATISTICS

Given its purpose and orientation, *Statistics for Lawyers* is less interested — and less effective — in examining legal doctrine and practice.³⁰ Just as the focus on legal applications makes *Statistics for Lawyers* a most unusual statistics text, so too, its bare bones approach to legal doctrine and cases takes it outside the mainstream of law books. *Statistics for Lawyers* is neither casebook nor hornbook.³¹ The courts and lawyers confronting statistical studies are visible only as skiagrams. The notes on legal doctrine, while accurate, are the soul of brevity, and the citations to the relevant cases and literature are not uniformly complete and current.³²

This limitation³³ is especially apparent when one considers specific topics in greater depth. I shall choose but one — the application of mathematics in a criminal case to express the probability of a disputed event indicative of guilt or innocence. Furthermore, of the many cases involving such “probability evidence,”³⁴ I shall analyze only one —

30. *Statistics for Lawyers* does not purport to investigate the problems arising from the injection of statistics and statistical experts into the legal process. A few recent books do focus on the practical problems of forensic statistics. The most significant is *Statistical Assessments supra* note 23, a National Research Council committee report containing thoughtful recommendations for improving the practice of attorneys and statisticians. Essays and case studies by statisticians involved in legal proceedings are collected in *STATISTICS AND THE LAW* (M. Degroot, S. Fienberg & J. Kadane eds. 1986). For reflections on these works and on the emerging specialty of “legal statistics,” see Kaye, *Improving Legal Statistics*, 24 *LAW & SOC. REV.* 1255 (1991); Kaye, *Statistical Proof in the Courts*, 35 *CONTEMP. PSYCH.* 839 (1990).

31. Cf. BARNES, *STATISTICS AS PROOF: FUNDAMENTALS OF QUANTITATIVE EVIDENCE* (1983) (casebook). Finkelstein and Levin explain that “[j]udicial opinions are not given because they generally do not elucidate the statistical issues that are our primary concern.” P. vii.

32. *Statistics for Lawyers* has a section on probabilities relating to “hair evidence” (pp. 88-91). A number of the cases cited there as apparently involving the probability of matching hair fibers involve no such evidence. *E.g.*, *State v. Kim*, 398 N.W.2d 544 (Minn. 1987); *People v. Riskey*, 214 N.Y. 75, 108 N.E. 200 (1915). They are, however, related to the broader question of the admissibility of “population frequency statistics.” Moreover, the references to the controversy in the forensic scientific literature omit not only the study, Wickenheiser & Hepworth, *Further Evaluation of Probabilities in Human Scalp Hair Comparisons*, 35 *J. FORENSIC SCI.* 1323 (1990), which appeared after the publication of *Statistics for Lawyers*, but also Aitken & Robertson, *A Contribution to the Discussion of Probabilities and Human Hair Comparisons*, 32 *J. FORENSIC SCI.* 684 (1987), Miller, *Procedural Bias in Forensic Science Examinations of Human Hair*, 11 *LAW & HUM. BEHAV.* 157 (1987), as well as all law review articles on the subject. Finkelstein and Levin also list most of the unique Minnesota cases rejecting well-founded probabilities (pp. 90-91), but fail to note that Minn. Stat. § 634.25 (Supp. 1989) repudiates these cases as applied to “statistical population frequency evidence” involving genetic markers. For other examples of significant literature not included in *Statistics for Lawyers*, see *supra* notes 4-11.

33. I use the term “limitation” descriptively rather than pejoratively. Finkelstein and Levin cannot be faulted for choosing to write a statistics text instead of a study of the law. Likewise, my citations to items published in 1990 are intended to supplement the references in *Statistics for Lawyers* rather than to criticize the authors for not mentioning material that was not available to them.

34. For examples, see McCORMICK ON EVIDENCE § 210 (J. Strong 4th ed. forthcoming 1991); Kaye, *The Admissibility of “Probability Evidence” in Criminal Trials* (pts. 1 & 2), 26 *JURIMETRICS J.* 343 (1986), 27 *JURIMETRICS J.* 160 (1987).

Branion v. Gramly.³⁵ *Statistics for Lawyers* makes no serious effort to describe the case or to investigate the general question of the admissibility of probability evidence,³⁶ but I shall show how it provides the basic tools for a more probing study.

As the Court of Appeals for the Seventh Circuit describes *Branion*, Donna Branion . . . was strangled and shot at least four times. She was not molested; there were no signs of forced entry into the apartment, from which nothing was stolen. This led the police to doubt that a stranger was responsible. A jury concluded that Donna's husband, John M. Branion, Jr., did the deed. The evidence was circumstantial, but what circumstances!³⁷

What circumstances, indeed. John and Donna Branion came from prominent Chicago families. John's father was a well-known attorney and the city's deputy chief public defender.³⁸ Donna's father was a wealthy banker.³⁹ John himself held "a position of responsibility"⁴⁰ as a physician at a hospital in Hyde Park, and had marched with Martin Luther King, Jr.⁴¹ The state judge who presided over the trial may have tried to collect money from John's many friends in exchange for a judgment of acquittal notwithstanding the verdict.⁴²

When appeals in the Illinois courts proved unavailing and the conviction became final, John fled, making his way to Uganda. There he allegedly became Idi Amin's physician until that dictator was overthrown in 1979.⁴³ The Ugandan government returned Branion to the United States in 1983, and he began serving his sentence of imprisonment. Three years later, he filed a motion for a writ of habeas corpus, asserting his innocence and maintaining that no rational jury hearing the state's evidence could have found him guilty. Two federal district judges ruled to the contrary, and Branion's appeals from these decisions came to the Seventh Circuit.

Branion's attorneys — three law professors at Northwestern Uni-

35. 855 F.2d 1256 (7th Cir. 1988).

36. It cites *Branion*, but only for the very general proposition that "statistical methods, properly employed, have substantial value. . . . Nothing about the nature of litigation in general, or the criminal process in particular, makes anathema of additional information, whether or not that knowledge has numbers attached" (p. 94).

37. 855 F.2d at 1256.

38. 855 F.2d at 1258. Raised by a white family, John Marshall Branion became Chicago's first black public defender. D'Amato, *The Ultimate Injustice: When a Court Deconstructs the Facts*, 11 CARDOZO L. REV. 1313, 1317 (1990).

39. 855 F.2d at 1258. D'Amato, *supra* note 38, at 1317, remarks that the family was "one of the wealthiest black families in the nation."

40. 855 F.2d at 1258.

41. D'Amato, *supra* note 38, at 1317-18, describes more of Branion's involvement in the civil rights movement, and Heise, *John Branion, recently freed in wife's slaying*, Chicago Tribune, Sept. 14, 1990, § 2, at 8, col. 5, reports that Branion was King's personal physician.

42. 855 F.2d at 1258-59. The judge was convicted of extortion in connection with other cases. 855 F.2d at 1258-59.

43. 855 F.2d at 1259 (citing New York Times and Associated Press accounts).

versity⁴⁴ — submitted a brief relying on a “standard deviation curve” to show that “the chance that Dr. Branion is guilty . . . is one in 9,000.”⁴⁵ The response was blistering.⁴⁶ Judge Easterbrook’s opinion for a unanimous panel ranges across sources as diverse as “Werner Heisenberg’s work”⁴⁷ and Isaac Todhunter’s 1865 *History of the Mathematical Theory of Probability*.⁴⁸ It may be the world’s first appellate court opinion to take the partial derivatives of a function.⁴⁹ And it is scathing in its criticism of counsel’s use of probability theory. The bemused reader of this judicial *tour de force* may wonder who is right: the attorney-academics, the academician turned judge,⁵⁰ or neither? To answer this question, we must begin with counsel’s computations.

A. Constructing a Probability Model

John Branion’s petition for a writ of habeas corpus relied on the defense of factual impossibility. He argued that he could not have murdered his wife within the relevant time-period, for even under the state’s theory of the crime, that would have required him to drive from the hospital (where he was working until 11:30) to his home (making two stops *en route*),⁵¹ to strangle and shoot his wife, to clean his hands, to dispose of his gun, and to summon his neighbors — all by 11:57 (when the police logged a call reporting the death). To show that he could not complete all these tasks in these twenty-seven minutes, Branion invoked some elementary principles of probability and

44. Professor Anthony D’Amato represented Branion at oral argument. 855 F.2d at 1257. With him on the brief were Professors Thomas F. Geraghty and Jon R. Waltz and two students working in the Northwestern University Legal Clinic.

45. 855 F.2d at 1270. Professor D’Amato has advised me that he alone bears responsibility for the probability argument. Letter from Anthony D’Amato to David Kaye (Nov. 10, 1989) [hereinafter D’Amato letter].

46. The court’s displeasure extended beyond the probability argument. See 855 F.2d at 1260-61.

47. 855 F.2d at 1263 n.5. Heisenberg’s uncertainty principle, to which the court alludes, states that certain pairs of physical variables, like the position and momentum of a particle, cannot be measured simultaneously with arbitrary precision. It is all too easy to misinterpret this fundamental feature of the subatomic world as stating that an observer always interferes with the system being observed. See, e.g., Tribe, *The Curvature of Constitutional Space: What Lawyers Can Learn from Modern Physics*, 103 HARV. L. REV. 1, 17-20 (1989); Tribe, *Law’s Geometry and the Curvature of Constitutional Space*, 44 REC. A. B. CITY N.Y. 575, 576 (1989). The uncertainty principle supports no such implication, and this feature of the microscopic realm of quantum mechanics has no real application to the macroscopic events that constitute the litigator’s world. In short, the uncertainty with which judges and jurors must deal is unrelated to “Heisenberg’s work.”

48. 855 F.2d at 1264.

49. 855 F.2d at 1265-66 & n.7.

50. Frank Easterbrook, a specialist in antitrust law and economics, came to the bench from the University of Chicago faculty.

51. Branion picked his son up from nursery school and made another stop to talk with a friend with whom he was to have had lunch. 855 F.2d at 1262.

statistics.⁵² Branion's brief to the Seventh Circuit argued that if the impossibly short time of six minutes were allowed for all the events other than pure driving and strangling, he still would have had to drive from the hospital and complete the strangling in twenty-one minutes — a feat that could not be reconciled with the evidence:

[T]here were two key pieces of uncontradicted prosecutorial evidence given in the form of a range of estimated times: Dr. Belmonte's evidence that the bruises on Donna Branion's neck took between 15 and 30 minutes to form, and Detective Boyle's evidence that the driving time of Dr. Branion's route took between 6 to 12 minutes. When evidentiary ranges are given, the most probable events can be plotted on a standard deviation curve or "bell curve":

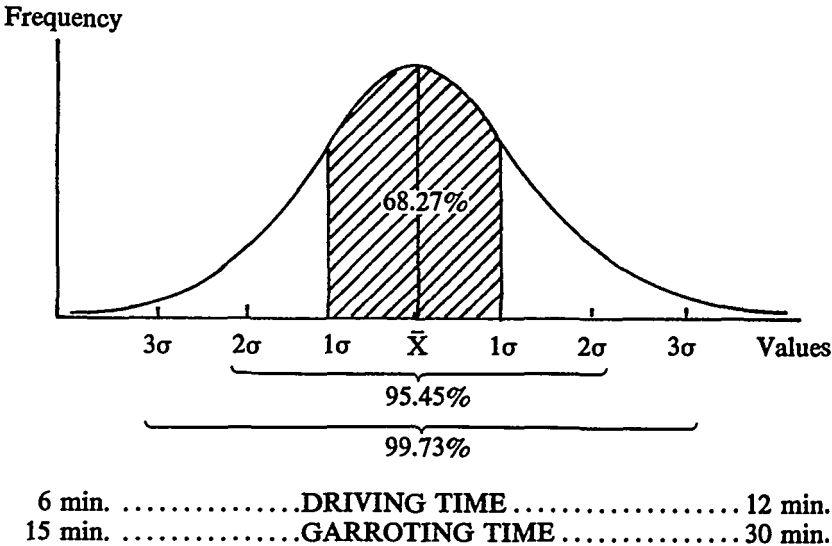


Figure 1

... At (a) 15 minutes and (b) 6 minutes, the probabilities are less than 1%. Moreover, since the two time-sequences are independent of each other, the probabilities must be multiplied to find the chances that both were applicable to Dr. Branion. Hence the probability that the time between garroting and shooting was 15 minutes *and* the probability that the driving time was 6 minutes is $.01 \times .01 = .0001$ A reasonable doubt probability of 90%, or 0.9, is 9,000 times larger in magnitude than the joint probability of .0001. In other words, if there were 9,000 cases similar to Dr. Branion's where the only two facts known were the garroting time and the driving time, then 8,999 defendants would be innocent and only one guilty. In short, the chance that Dr. Branion is guilty, on

52. Branion also adduced more traditional arguments about the timing of the murder. For example, he emphasized that a neighbor had heard "three sounds and the sounds of some commotion" from Branion's house before Branion had even left the hospital. Brief for Petitioner-Appellant at 18.

the basis of the garroting time and the driving time alone, is one in 9,000.⁵³

As previously indicated, the 1/9000 figure for the probability of Branion's guilt did not impress the court, at least not favorably. First, noticing that the "jury could have found that 30 minutes lapsed between Branion's leaving the Hospital and his call to the police," the court concluded that "[w]e therefore should like to know the probability that the combination of travel and murder times came to 30 minutes or less" rather than the twenty-one minutes or less discussed in Branion's brief.⁵⁴

This criticism suggests that Branion set out to solve the wrong problem, but it does not challenge the method he used to attach a probability to the combined time being less than some critical figure. The appropriate number may be twenty-one minutes, thirty minutes, or some other value, but this issue need not detain us. In the face of unavoidable uncertainty in the times for miscellaneous acts other than driving and garroting, the best procedure would be to compute the pertinent probabilities for all the plausible numbers in the twenty-one to thirty minute range.

The more fundamental question is the manner in which the probability for any of these times should be computed, and the opinion attacks the methodology of the *Branion* brief on several grounds. To analyze these statistical issues clearly, it is essential to specify the probabilistic model that underlies Branion's computation — an undertaking left unfinished by both Branion and the court, but one that *Statistics for Lawyers* can help complete.⁵⁵ We want to know how long it would take Branion to drive to his house and to strangle a person of a certain bone structure, leaving a bruise of a given depth without breaking any bones.⁵⁶ To measure this quantity, we could conduct the following experiment: have people of Branion's skill at driving and strangling drive the route and strangle similar victims to the same degree of bruising. This hypothetical experiment would generate a distribution, that is, a list a relative frequencies for each value of the total elapsed time Z . If nothing changed from one instance to another, Z would be constant. We could express this idealized situation with a monumentally simple equation:

$$Z = \text{CONSTANT.} \quad (1)$$

This equation merely states that the same time would be recorded for each trial. The relative frequency of the times, $f(Z)$, would have the

53. 855 F.2d 1256 app. at 1270 (emphasis in original).

54. 855 F.2d at 1265.

55. For the book's expositions of probability distributions, see pp. 23-61, 111-55.

56. "Donna Branion was strangled as well as shot. Her neck was bruised; she had been garrotted with the cord from her iron. The pressure was not great enough to break any bones." 855 F.2d at 1262.

value one whenever Z is the observed constant, and the value zero for all other values of Z .

Of course, our horrid experiment would not have this trivial an outcome. Things can and will change from one instance to another.⁵⁷ On days when there is little traffic, when the driver is lucky enough to meet only green stoplights, when the victim can be garroted with a minimum of fuss, and when a bruise will form quickly, Z will be small. When traffic and road conditions are less favorable and the victim less conducive to prompt garroting and bruising, Z will be large. Many small, effectively random factors would influence the magnitude of Z , and we could represent the combined effect of these variables with a revised equation:

$$Z = \text{CONSTANT} + \delta. \quad (2)$$

According to this model, the measured times Z are just the sum of the previous constant term and an "error" or "disturbance" term δ , reflecting the impact of the many things that cause Z to vary from one test to the next. Sometimes the net effect is to lengthen Z , sometimes to shorten it. The average of these disturbances, however, may be taken to be zero,⁵⁸ so that the mean value of Z is just the constant term. With these assumptions, the distribution $f(Z)$ is no longer a solitary, sharp spike; it covers a broader band of values centered at CONSTANT .

If we really knew the constant and the error function, we would have no problem finding the probability that the time would be less than any given amount z . The probability is the area under the distribution $f(Z)$ to the left of z , as sketched in Figure 2, which arbitrarily (at this point) displays a normal distribution.

Obviously, we cannot perform this experiment, and Branion certainly did not examine the distribution $f(Z)$ to deduce the 1/9000 probability that he paraded before the court. Instead, he (1) considered two different distributions, (2) made some strong assumptions about them, (3) looked at the area in each of their tails, and (4) multiplied these areas together. This first step is fine, at least in principle. The second is more defensible than the court acknowledged. The third and fourth are blunders.

57. Ideally, we would want to ensure the relevant conditions during the experiment matched those on the day in question. But perfect control is impossible and some influences may be overlooked. The result even with a suitably controlled experiment is variability in the measured times.

58. The error term can be defined to have a mean of zero without affecting the generality of the model. Suppose that an error term (call it ϵ) had a nonzero mean m . Then by defining δ to be $\epsilon - m$, we obtain a transformed error term with mean zero. The equation now reads

$$Z = (\text{CONSTANT} + m) + \delta$$

which has the desired form, a constant term (equal to the old constant term plus m) plus an error term with mean zero.

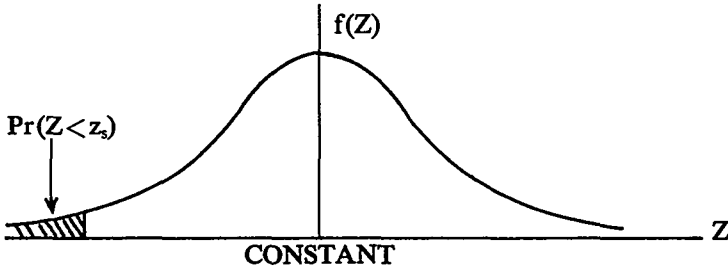


Figure 2

One possible distribution of times reflecting “smearing” due to random variables, and the probability of an observed time being less than some number z_c .

Because the total time Z cannot be measured directly, it is reasonable to focus on some components that can be. In other words, one can consider $Z = X + Y$, where X is the driving time and Y the bruising time, try to ascertain their distributions $g(X)$ and $h(Y)$ separately, and then combine them to find $f(Z)$. At the outset, it is easy enough to write down general equations that might describe X and Y . A model such as equation (2) seems appropriate. That is, we assume that

$$X = \text{CONSTANT}_X + \delta_X \tag{3}$$

and that

$$Y = \text{CONSTANT}_Y + \delta_Y \tag{4}$$

where CONSTANT_X and CONSTANT_Y are constants, and δ_X and δ_Y are error terms.

Because the error terms can be defined so that they average out to zero,⁵⁹ the constants can be estimated by the means of the observed times. The police drove the route six times in 1968, with results ranging from six to twelve minutes.⁶⁰ The mean time for the sample is the arithmetic average of these six measurements. Despite the fact that virtually any statistician would use the sample mean to estimate the constant term, Branion did not report this number. Instead, he averaged only the two extreme measurements.⁶¹ Still, this number could be close to the unreported sample mean.⁶²

Likewise, Branion averaged the end points of the pathologist’s esti-

59. See *supra* note 58.

60. 855 F.2d at 1265.

61. 855 F.2d at 1265.

62. Trial counsel never asked the police officers for any of the driving times other than the two extremes. Trial transcript at 270-93.

mate of fifteen to thirty minutes for the time needed for a bruise to form.⁶³ Nevertheless, the court's rhetoric that "we have no idea what the mean . . . might be" seems overdrawn. If we credit the testimony, then we know that the mean is more than fifteen minutes and less than thirty, and if we presume that people giving a range of possibilities usually have in mind a central point plus or minus some quantity, then we do have an idea that the mean is near the midpoint of the fifteen to thirty minute interval, as Branion posited.

Despite the doubts about Branion's numbers, let us proceed as if they are good estimates of the constants in (3) and (4). We still need to specify the error terms, for they capture the randomness in the times X and Y . Branion assumed, in effect, that the error terms were normally distributed⁶⁴ with standard deviations one and 2.5 minutes, respectively.⁶⁵

The court dismissed these assumptions out of hand. Of the postulated distribution of driving times, it said, "Nothing suggests a Gaussian distribution."⁶⁶ Of the postulated distribution of bruising times, it remarked, "[W]e have no idea what the mean time or standard deviation might be."⁶⁷ The court did not explicitly denounce the assumption of normality for the bruising time, but it probably meant to.

Yet, the suggestion of normality is not so outlandish. To be sure, the testimony itself, involving the range for only six driving times and an expert's guess as to the range of bruising times, is dreadfully limited. But to observe that this evidence alone does not suggest a normal distribution is not to demonstrate that the assumption is unreasonable. Many physical and biological processes give rise to normal distributions. A remarkable theorem in statistics helps explain the prevalence of these distributions. This Central Limit Theorem, as it is called, asserts that the sum of a large number of random variables has an approximately normal distribution *even when the underlying variables are not normal*.⁶⁸ Now the disturbance terms in equations (3) and (4) represent the combined effect of many essentially random factors, such

63. The testimony of the pathologist in this regard is slightly ambiguous. See Trial transcript at 470-72.

64. A normal distribution is a relative frequency curve (more precisely, a probability density function) with some highly specific characteristics. It is one of a family of curves that are, as Branion's brief put it, "bell-shaped." 855 F.2d 1256 app. at 1270; see pp. 25, 117-18.

65. The standard deviations measure the extent of the variations in the times that would be obtained in repeated experiments. See p. 42.

66. Branion v. Gramly, 855 F.2d 1256, 1265 (7th Cir. 1988). Normal distributions are often called Gaussian, especially in the physical sciences, in honor of Karl Frederick Gauss' work on error distributions (p. 119); see also 1 THE PROBABILISTIC REVOLUTION 272 (L. Kruger, L.J. Datson & M. Heidelberger eds. 1987).

67. 855 F.2d at 1265. The court disparaged the pathologist's statement of the range of times as "a number from the air." 855 F.2d at 1265.

68. See pp. 118-20; D. MOORE & G. McCABE, INTRODUCTION TO THE PRACTICE OF STATISTICS 416 (1989).

as, for (3), the timing of stoplights, how many cars are at each intersection, and so on. The resulting delays are effectively random variables, and the theorem therefore implies that the scatter of observed times about the constant terms could well be approximately normal. Although the court is correct to say that this normal model cannot be derived or verified from the limited data, one could still maintain that the data are not inconsistent with this theoretical picture.

Suppose, then, that we indulge the assumption of normality. We accept equations (3) and (4) with the added conditions that each error term is normally distributed. Two parameters, the mean and the standard deviation, are sufficient to specify a normal distribution. The mean gives the location of the center of the distribution, and the standard deviation indicates the width of the curve, as shown in Branion's brief.⁶⁹ Because the constant terms express the regularity in the times, the error terms must have means of zero.⁷⁰

Branion's estimates for the standard deviations are difficult to defend. Ordinarily, one would estimate the standard deviation of the error term with a sample standard deviation.⁷¹ The latter quantity is easily computed.⁷² Instead of calculating the sample standard deviation, however, Branion blithely took one minute as the standard deviation. Even without knowing the four intermediate driving times in the six to twelve minute interval, the usual estimator of the standard deviation must have a larger value. A little arithmetic reveals that it can be no less than 1.9 minutes.⁷³ Although the court did not point out that Branion's estimate for the standard deviation was mathematically impossible, it did sense an incongruity between the data and Branion's wildly optimistic figure.⁷⁴

69. See *supra* text accompanying note 53.

70. See *supra* note 58.

71. When the sample standard deviation is used, it may be better to use a Student's curve rather than a normal curve to compute the tail-end probability. See, e.g., pp. 224-25; D. FREEDMAN, R. PISANI & R. PURVES, *STATISTICS* 463 (1980). This would increase the probability by an amount that depends on the size of the sample used in estimating the standard deviation.

72. For the six measurements of the driving time, which we may designate X_1, X_2, \dots, X_6 , the sample standard deviation is $[\{(X_1 - M)^2 + (X_2 - M)^2 + \dots + (X_6 - M)^2\}/5]$, where M is the sample mean. One might ask why the sum of the squared deviations from the mean is divided by five instead of six. After all, there were six observations. It turns out that dividing by the number of sample observations produces a biased estimator, and dividing by the number of observations less one corrects this problem.

73. The smaller the standard deviation, the closer the measurements tend to cluster around the mean. Knowing the most extreme values $X_1 = 6$ and $X_2 = 12$, we can deduce the range of possible sample standard deviations. The smallest standard deviation occurs when the unreported measurements each equal the sample mean, for in that situation these observations contribute nothing to the sum of the squares. Thus, $\min[S^2] = [(6-9)^2 + (12-9)^2]/5 = 3.6$, and $\min[S] = \sqrt{3.6} = 1.9$. The largest sample standard deviation occurs when the unreported values are as far from the sample mean as possible — when they are at the extremes. Thus, $\max[S^2] = [3(6-9)^2 + 3(12-9)^2]/5 = 3\min[S^2] = 10.8$, and $\max[S] = \sqrt{10.8} = 3.3$.

74. The opinion noted that a standard deviation of one minute would imply that the six minute time was three standard deviations from the mean of nine minutes, and that deviations

Oddly, Branion's estimate for the standard deviation of the bruising time, which the court disdainfully dismissed as "a number from the air,"⁷⁵ may be more defensible. Perhaps pathologists, when they toss out a range of times, do think in terms of two or three standard deviations.⁷⁶ I, for one, tend to doubt this, but Branion's choice of a standard deviation for the bruising time does not contradict the evidence quite so flagrantly.

Giving the benefit of all doubts to Branion, suppose we take the driving time X to be normally distributed with mean 9 and standard deviation 1.9, and the bruising time to be similarly distributed with mean 22.5 and standard deviation 2.5. Where does this get us? Branion's answer was to use such quantities to sketch the resulting distributions of X and Y , and to multiply the two tail-end probabilities because "the two time-sequences are independent of each other."⁷⁷ This the court characterized as "mindless multiplication"⁷⁸ produced "by a method that is proper only if the probability of the 6-minute drive and the 15-minute bruise are independent events,"⁷⁹ when "on the state's hypothesis of a planned murder, they are anything but independent."⁸⁰

Whether the times are dependent, however, is not so clear, and even if they are, the true villain may not be the claim of independence. The rules for combining probabilities also justify the multiplication of probabilities of dependent events.⁸¹ To analyze the assumption of statistical independence and to show that this assumption is not essential to Branion's computation, we must clarify the meaning of independence. Then we turn to a flaw in that calculation that cannot be explained away — the specification of the event whose probability we wish to calculate.

Neither the court nor counsel explained what "independence"

this large or larger should arise much less frequently than the one out of six times reported by the police if the distribution really were normal. *Branion v. Gramly*, 855 F.2d 1256, 1265 (7th Cir. 1988). Of course, this point applies to the twelve minute time as well. Although Branion's figures implied that values as extreme as the two he reported would occur very rarely, they evidently had been observed twice in the sample of six.

75. 855 F.2d at 1265.

76. I have assumed that the pathologist is giving a range for a distribution of objectively known bruising times. Some statisticians would prefer to say that the pathologist is describing a subjective, or personal distribution that measures his degree of belief in each possible bruising time. They would ask whether he would agree that his distribution is approximately normal and has the standard deviation claimed by Branion. Even if there are no specific data on bruising times known to the pathologist, they might be willing to proceed on the basis of the expert's subjective judgment.

77. 855 F.2d at 1270.

78. 855 F.2d at 1264.

79. 855 F.2d at 1265.

80. 855 F.2d at 1265.

81. Pp. 81, 87; cf. Fairley & Mosteller, *A Conversation About Collins*, 41 U. CHI. L. REV. 242 (1974) (implications for *People v. Collins*).

means here, and the concept is surprisingly subtle.⁸² Statistical independence does not mean causal independence.⁸³ It simply means that one event is neither more nor less likely to occur when another one does. If A is one event and B another, then A is independent of B if and only if the conditional probability of A given B equals the unconditional probability of A , that is, if $Pr(A|B) = Pr(A)$ (p. 81). When A and B are independent in this sense, knowing that B has occurred should not affect our judgment of the probability that A will occur.

Branion reasoned that "since the two time-sequences are independent of each other, the probabilities must be multiplied to find the chances that both were applicable to Dr. Branion."⁸⁴ If the driving time X is independent of the bruising time Y , then it would indeed follow from the definition of conditional probability⁸⁵ that the probability of both a small driving time and a small bruising time is the product of the probability of each of these small times:

$$Pr[(X \leq x_c) \& (Y \leq y_c)] = Pr(X \leq x_c)Pr(Y \leq y_c). \quad (5)$$

Branion took the individual probabilities to be 0.01, yielding a joint probability of 0.0001. Because the probability associated with values as small as three standard deviations below the mean of a normal curve is 0.0013,⁸⁶ however, the joint probability under Branion's view of the evidence is actually much less. It is $(0.0013)^2 = 0.0000017$. On the other hand, we also have seen that Branion erred in inferring or positing a standard deviation of driving times of only one minute.

82. See, e.g., Kruskal, *Miracles and Statistics: The Casual Assumption of Independence*, 83 J. AM. STATISTICAL A. 929 (1988); Meier & Zabell, *Benjamin Pierce and The Howland Will*, 75 J. AM. STATISTICAL A. 497, 501-02 (1980). *Statistics for Lawyers* contains no sustained analysis of these subtleties.

83. Two events A and B may be dependent even when A does not influence B and vice versa. Thus, the number of gallons of ice cream consumed per year may rise along with the number of crimes committed by juveniles per year (because of a growing juvenile population). Although eating ice cream does not cause crime, and crime does not cause people to eat ice cream, the quantities of ice cream consumption and juvenile crime nevertheless would not be statistically independent. Cf. G. YULE & M. KENDALL, *AN INTRODUCTION TO THE THEORY OF STATISTICS* 315-16 (14th ed. 1950) (nearly perfect correlation between number of radios and number of mental defectives per 10,000 people in United Kingdom from 1924 to 1937).

84. Branion v. Gramly, 855 F.2d at 1270.

85. The probability of A conditioned on B is the joint probability divided by the probability of B . In symbols, $Pr(A|B) = Pr(A \text{ and } B)/Pr(B)$. To say that A and B are independent, however, means that $Pr(A|B) = Pr(A)$. Substituting $Pr(A)$ for $Pr(A|B)$ yields $Pr(A) = Pr(A \text{ and } B)/Pr(B)$. Multiplying both sides of this last equation by $Pr(B)$ gives the result that $Pr(A)Pr(B) = Pr(A \text{ and } B)$. In other words, when A and B are independent, the probability of both A and B is the probability of A times the probability of B . In fact, it is common to take this multiplication rule as defining independence and to deduce that for independent events A and B , $Pr(A|B) = Pr(A)$. Whichever route one takes, however, the special multiplication rule described here holds if and only if A and B are independent.

86. Why the brief would advance the 0.01 figure is a mystery. That the true value is much smaller, and therefore more supportive of Branion's case, is apparent from the picture in Branion's brief. As depicted there, 99.73% of the area under the normal curve lies within plus-or-minus three standard deviations. Hence, .27% lies in the two tails, and half of 0.27%, or 0.135% = 0.00135, is in each tail.

Since the smallest standard deviation consistent with the data on driving times is 1.9 minutes, equation (5) leads to the conclusion that the probability that $X \leq 6$ and $Y \leq 15$ is $(0.057)(0.0013) = 0.00007$.⁸⁷ On the assumptions most favorable to Branion, then, the joint probability is even smaller than Branion and his counsel imagined.

But are short driving times and short bruising times uncorrelated, as this analysis supposes? We want the times Y for people who know how long it took them to drive the route. If these people can and would inflict a bruise more quickly to compensate for a large X , then Y is not independent of X . If they might be more leisurely in the strangling when the drive went quickly, then, once again, Y is not independent of X . On the other hand, if the "garroting time" depicted in the brief is predominantly the time it takes for a bruise to form from blood flowing from broken capillaries, then even a murderer intent on minimizing the "garroting time" may be unable to exert much control over Y . Because the trial testimony sheds little light on the reasons for the variability in Y , the court's criticism of the independence assumption must be justified on the ground that a reviewing court should examine the evidence in the light most favorable to the prevailing party. Without a record or judicially noticeable facts indicating independence, the appellate court may not wish to entertain so favorable an assumption for the proponent of the probability analysis.

Still, joint probabilities exist for dependent variables, just as they do for independent ones. The only difference is that some knowledge of the relationship between the variables is required to find the probability that both are small. Rather than considering X and Y separately, we can consider the probability of each possible pair of values X and Y . This way of thinking leads to a joint probability distribution $p(X, Y)$ from which the probability that X and Y are within a given region is easily obtained.⁸⁸ Of course, we have already seen that using the limited evidence in *Branion* to specify the unconditional distributions $f(X)$ and $g(Y)$ of the driving and bruising times is difficult enough. Using it to obtain a joint distribution that assigns probabili-

87. These numbers are the area in the left-hand tails of the pertinent normal curves. See *supra* Figure 1. The value $X = 6$ is $(9 - 6)/1.9 = 1.58$ standard deviations below the mean of 9 for X . The probability of a value at least this far below the mean for a normal random variable is 0.057. Likewise, the value $Y = 15$ is $(22.5 - 15)/2.5 = 3$ standard deviations to the left of the mean for Y . The probability for a Y this small or smaller for a normal random variable is about .0013.

88. The joint probability of X not exceeding x , and Y not exceeding y , is the volume under the joint distribution and above the rectangle whose corners are $(0,0)$, $(x,0)$, $(0,y)$ and (x,y) :

$$\Pr(X \leq x, Y \leq y) = \int_0^x d\tau \int_0^y p(X=\tau, Y=\phi) d\phi.$$

This definition is not presented in *Statistics for Lawyers*, but the text does speak of "two variables with a joint distribution" (p. 51).

ties to all the possible combinations of these times seems out of the question.

There is, however, a way to deduce the probability of a joint event that does not demand the full joint distribution, but only certain slices through it. Rather than worrying about all possible combinations of X and Y , one can consider the distribution of possible bruising times Y given specific values for the driving time X . In estimating the range of bruising times, the pathologist may well have known that the state's theory involved a very brief driving time. Arguably, his estimate for Y was conditioned on the assumption of a small value x_r of X . If so, he was describing not the unconditional distribution $g(Y)$, but a conditional distribution $g(Y|X = x_r)$.⁸⁹ If we indulge the further assumption that this conditional distribution is approximately the same for all small driving times and if we interpret the pathologist's meager remarks as positing a normal distribution with a well-specified mean and standard deviation, then the probability of particular values of X and Y is a product akin to one that Branion proposed. Analogous to equation (5), we might write:

$$\Pr[(X \leq x_r) \text{ and } (Y \leq y_s)] = \Pr(X \leq x_r) \Pr[(Y \leq y_s)|(X \leq x_r)] \quad (6)$$

In other words, the joint probability that the driving time will not exceed x_r minutes and that the bruising time will not exceed y_s minutes is (a) the probability that the driving time will not exceed x_r , multiplied by (b) the probability that the bruising time will not exceed y_s , given that the driving time does not exceed x_r .

This reinterpretation of the pathologist's testimony sidesteps the court's complaint that the events are "anything but independent."⁹⁰ It acknowledges their dependence and purports to account for it by using the multiplication rule that applies to dependent events. In this way, it seems to justify the conclusion that the evidence about driving and bruising times indicates a probability for $X \leq 6$ and $Y \leq 15$ of 0.00007. But even if the problem of independence can be surmounted in this fashion, an additional objection to Branion's analysis remains.

The court recognized that more was at work than the dubious independence assumption, for it stated:

That's *still* not all. Even if the time sequences are independent, even if we are interested in the probability that the driving plus choking time is 21 minutes or less, even if the distributions are Gaussian, the

89. Cf. p. 51 (referring to "the conditional distribution of Y given fixed values of $X = x$ "). This interpretation of the pathologist's testimony came to mind as a result of remarks by Professor D'Amato. Letter from Anthony D'Amato to David Kaye (Sept. 22, 1989) ("any probabilistic inference you make concerning the killer's time-dependence at the lowest end of the spectrum is wholly offset by the fact that loading the spectrums at the minimum time is already built into the state's case because the state has the burden of proof"). Professor D'Amato does not subscribe to my formulation, which he complains "[does not do] justice to my arguments contained in our correspondence." D'Amato letter, *supra* note 45.

90. Branion v. Gramly, 855 F.2d 1256, 1265 (7th Cir. 1988).

probability is very sensitive to the assumed standard deviation. On Branion's assumptions, the probability is 0.1% rather than 0.01% as Branion believes; on more plausible assumptions, the probability is 10%.⁹¹

An accompanying footnote explains that "[w]e need to compute, from the means and variances of two or more independent events, the joint mean and variance."⁹² Resorting to a general formula for the variance of any differentiable function of N independent variables, and using two distinct pairs of values for the standard deviations of the driving and bruising times, the court arrives at the probabilities of 0.001 and 0.10.⁹³

The court's discussion, although basically correct, is marred by computational errors and is needlessly opaque. It can be clarified, not by any particular formula in *Statistics for Lawyers*, but by a careful articulation of the problem. Recall that we introduced the separate distributions X and Y for the driving and bruising times only because the total time Z could be expressed as their sum. The event of interest is whether or not Z is less than or equal to the minimum time that Branion would have needed to commit the murder. Branion's brief, however, computed the probability that the driving time X would not exceed six minutes and that the bruising time Y would not exceed fifteen minutes. The court stated that a different procedure must be used to find the probability that $Z \leq 21$, but it never explained why. Apparently, the authors of the *Branion* brief believed that this range of outcomes for X and Y coincided with the event that the sum of these times would not exceed twenty-one minutes. Since $X \leq 6$ and $Y \leq 15$ imply that $X + Y \leq 21$, this implicit assumption may seem justified, and the court's opinion, which was otherwise quite detailed, did not explicitly question it. However, the premise is fallacious. A multiplication procedure like that in equation (6) is inappropriate for the simple reason that $X + Y \leq 21$ does not imply that $X \leq 6$ and $Y \leq 15$. There are many ways to make the driving and bruising times add to no more than twenty-one minutes while having the driving time exceed six minutes or the bruising time exceed fifteen minutes.⁹⁴ Because the probability given by equation (6) neglects such possibilities, Branion's procedure of multiplying the two tail-end probabilities necessarily understates the probability of the event that is of interest — that $Z \leq 21$.

To find the probability of this event, we do not need to evaluate the partial derivatives of Z with respect to X and Y , as the court did. The necessary formula can be derived with only high school algebra.⁹⁵ The

91. 855 F.2d at 1265.

92. 855 F.2d at 1265 n.7.

93. 855 F.2d at 1266 n.7.

94. For example, let $X = 7$ and $Y = 14$. Then $X + Y \leq 21$ even though $X > 6$.

95. M. DEGROOT, *PROBABILITY AND STATISTICS* 159 (1975) (Theorem 4). This result is a special case of the more general formula given in the court's opinion.

result, as reported in *Statistics for Lawyers*, is that “[t]he variance [which is the square of the standard deviation] of the sum of two *independent* random variables is the sum of the separate variance” (p. 43):

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y). \quad (7)$$

Using Branion’s values of 1 and 2.5 for the standard deviations of X and Y , one concludes that the variance of Z is 7.25, corresponding to a standard deviation for Z of $\sigma(Z) = 2.69$.⁹⁶ Because the sum of two normal distributions is also normal, and its mean is just the sum of the means of each distribution,⁹⁷ the event of interest has a probability given by the area beneath the appropriate normal curve and to the left of $Z = 21$ minutes.⁹⁸ Thus, the probability of a sufficiently small time, using Branion’s impossible values for the parameters of the driving and bruising times, is below 0.0001.⁹⁹

There is irony in these numbers. The court’s effort to improve on Branion’s calculation, although correct in principle, inexplicably overstates the joint probability by a factor of twenty.¹⁰⁰ At the same time, Branion’s conceptual and arithmetic errors fortuitously cancel. The joint probability under a model that assumes independence of two normal random variables is about 0.1%, as Branion found by applying the wrong formula to the wrong numbers, and not 0.01% as the court found by misapplying the right formula. With the more plausible use of the smallest possible standard deviation of the driving time,¹⁰¹ the probability is about 0.0004,¹⁰² which is an order of magnitude larger than the 0.00007 figure obtained from Branion’s procedure of multiplying two tail-end probabilities.

In short, the court’s final criticism of Branion’s “mindless multiplication” — that it fails to combine properly the two random variables — is apt.¹⁰³ That procedure is wrong because it misstates the event whose probability is of interest, and thereby ignores outcomes that contribute to the probability that the total time is small. On the other hand, however egregious the error may be at the conceptual level, in

96. We have $\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y) = 1^2 + 2.5^2 = 1 + 6.25 = 7.25$, and $\sigma(Z) = \sqrt{7.25} = 2.69$. The court found the standard deviation of Z to be 2.73. 855 F.2d at 1266 n.7.

97. *E.g.*, M. DEGROOT, *supra* note 95, at 151 (Theorem 3).

98. This curve has mean $M(Z) = M(X) + M(Y) = 9 + 22.5 = 31.5$ minutes and standard deviation 2.69 minutes.

99. $Z = 21$ is $10.5/2.69 = 3.9$ standard deviations below the mean. The area in the corresponding left-hand tail of a standard normal curve is 0.000048.

100. The court’s figure of 0.1% divided by $\Pr(Z \leq 21)$ when $M(Z) = 31.5$ and $\sigma(Z) = 2.69$ is $0.00100/0.000048 = 20.8$ times too large.

101. *See supra* note 73.

102. Again, we have a normal distribution with mean 31.5 minutes. The variance of Z , however, is now $1.9^2 + 2.5^2 = 9.85$. The corresponding standard deviation is $\sigma(Z) = 3.14$. The area to the left of a point $10.5/3.14 = 3.35$ standard deviations below the mean is about 0.0004.

103. Branion v. Gramly, 855 F.2d 1256, 1264 (7th Cir. 1988).

this instance it has little practical import. A probability of one out of ten thousand or even four out of ten thousand is still awfully small.

So the question remains: How could a jury applying the beyond-a-reasonable-doubt standard rationally conclude that so improbable an outcome had occurred? One possibility is that the other evidence of guilt was extremely powerful, so that it would be reasonable to conclude that the improbable occurred.¹⁰⁴ To enhance the plausibility of this result, the court's opinion portrays the circumstantial evidence in a dramatic light.¹⁰⁵

In addition, the opinion shows that if the standard deviations of X and Y were much larger than Branion's estimates, the probability would rise to 0.1.¹⁰⁶ In this range, the probability is more easily outweighed by the state's evidence of guilt. In view of the terrible uncertainty in the estimates of the parameters on which the probability depends, this kind of sensitivity analysis is admirable. If the plausible variations in the parameter estimates had an insubstantial effect on the relevant probability, then some of the court's complaints would not have seriously undercut Branion's claim. Because the probability floated by Branion turns out to be sensitive to plausible changes in the values of the parameters, the 0.0001 figure cannot be decisive.

B. *Putting Probability in Its Place*

En route to its withering critique of Branion's computations, the court of appeals praised the "substantial value" of "[s]tatistical methods, properly employed."¹⁰⁷ Because the court was confronted only with a probability argument in an appeal on the sufficiency of the evidence, it did not consider the potential value of trial testimony or argument about probabilities. Indeed, there was no such testimony in *Branion*. Nonetheless, the court of appeals took the opportunity to

104. A less obvious possibility is that the probability that anyone in Branion's position could commit the crime in the available time really is very small — but the probability of each alternative explanation is smaller still. In this situation, even an "outrageous" possibility can emerge victorious. Meier & Zabell, *supra* note 82, at 502.

105. The court notes (1) the presence of a mistress whom Branion married shortly after the murder, (2) a Walther PPK and four shells of ammunition for it missing from Branion's gun collection, where ballistics experts determined that the murder weapon was a Walther PPK, and four shell casings were found near the body, and (3) some peculiar explanations that Branion offered the police. 855 F.2d at 1258. For the view that the opinion disingenuously overstates the significance of these items, see D'Amato, *Self-Regulation of Judicial Misconduct Could Be Mis-Regulation*, 89 MICH. L. REV. 609, 620-21 (1990); D'Amato, *supra* note 38.

106. The court wrote that "[i]f . . . the standard deviation of the driving times = 3 and the standard deviation of bruise-forming times is 7.5 minutes, . . . we quickly compute $[\sigma(Z)] = 8.1$ minutes, so that the probability that $[Z] \leq 21 \approx 0.10$." 855 F.2d at 1266 n.7. This computation is correct: $\sigma^2(Z) = 3^2 + 7.5^2 = 65.25$, and $\sigma(Z) = 8.08$; $Z = 21$ therefore is $(21 - 31.5)/8.08 = 1.3$ standard deviations from the mean, with corresponding probability $\Pr(Z \leq 21) = 0.097$. The court opines that these values for the standard deviations are "more plausible." 855 F.2d at 1265. The value $\sigma(X) = 3$, however, is near the maximum possible sample standard deviation. See *supra* note 73.

107. 855 F.2d at 1263-64.

comment on probability analysis at trials, and it equated "statistical inferences" with more traditional evidence.¹⁰⁸ Thus, the court opined, the lesson of previous cases of probability evidence "is not that statistical methods are suspect, but that people must be sure of what they are looking for, and how they can prove it before they start fooling with algebra."¹⁰⁹

Statistics for Lawyers is not wrong to cite *Branion* for these general propositions, but there is more to be learned from the case than this. In the sense described by Hume and other philosophers, all evidence is, at bottom, probabilistic, but probability theory may be applied to courtroom proof in at least two distinct ways. It may function as evidence in itself, or it may serve as a cognitive aid, helping the trier of fact to weigh other evidence. These distinct functions have different implications for the usefulness of probability theory, although both applications of the theory are potentially legitimate and desirable.

The use of probability as evidence occurs when some objective quantity, such as life expectancy,¹¹⁰ the probability of recovering from a disease,¹¹¹ or the chance of a black serving on a grand jury,¹¹² is probative of some material issue. In these cases, a well-defined chance model is both available for and essential to evaluating this quantity, and the meaning of the quantity is not especially controversial. As such, statistical methods directed at these quantities should not be especially suspect. As the *Branion* court announced, the only real concern in the probability-as-evidence cases is the soundness of the methodology.

The probabilistic reasoning in *Branion* is different. Counsel relied on probability theory not as an item of evidence in itself, but as a rhetorical device to suggest the impact that other evidence should have. The argument was that because the probability of the times falling within the window of opportunity open to John Branion was 1/10,000, no rational jury could have found him guilty. Such reliance on probability theory to tell jurors and judges what to believe about evidence has provoked acrimonious controversy, in part because it is more difficult to agree upon an underlying probability model for the evidence and also because not all probabilists agree that probability

108. 855 F.2d at 1264. The court stated:

Much of the evidence we think of as most reliable is just a compendium of statistical inferences. . . . Nothing about the nature of litigation in general, or the criminal process in particular, makes anathema of additional information, whether or not that knowledge has numbers attached. After all, even eyewitnesses are testifying only to probabilities

See also *DePass v. United States*, 721 F.2d 203, 207 (7th Cir. 1983) (Posner, J., dissenting) ("[M]ost knowledge, and almost all legal evidence, is probabilistic.").

109. 855 F.2d at 1264.

110. See *DePass v. United States*, 721 F.2d 203 (7th Cir. 1983) (Posner, J., dissenting).

111. E.g., Annotation, *Medical Malpractice: Measure and Elements of Damages in Actions Based on Loss of Chance*, 81 A.L.R. 4th 485 (1990).

112. E.g., *Vasquez v. Hillery*, 474 U.S. 254 (1986).

theory can model belief.¹¹³ While I do not believe that these considerations should preclude the use of probability theory as a cognitive aid,¹¹⁴ one can maintain that even well-crafted probability arguments about the significance of evidence should be suspect — not to the extent of dismissing them out of hand, but to ensure that these arguments are taken as a guide to thought and not a substitute for it.

Furthermore, in considering probability theory as a cognitive aid in the assessment of evidence, one must attend to another possible distinction not apparent in *Branion's* undifferentiated discussion of probability evidence. In some instances, such as proof based on blood antigens in rape and paternity litigation, the factfinder cannot possibly understand the probative value of the evidence without some statistical analysis.¹¹⁵ Therefore, almost all courts allow relevant quantitative descriptions in this situation.¹¹⁶ But it seems hasty to cite such cases, as the *Branion* court does,¹¹⁷ for the acceptability of explicit statistical analysis in instances where the trier of fact can appreciate the evidence without any quantitative testimony.¹¹⁸

Although case law provides little guidance, even here probability demonstrations should not be categorically excluded. The nature of the evidence may not compel the introduction of probability theory, and such demonstrations can never be binding, but they can be instructive: "The outcome of a competent model can be compared with intuition. If the two agree, that is a source of confidence; if they don't — that is a point of departure for critical appraisal."¹¹⁹

To be sure, in most cases even this weak use of probability theory may not be justified, especially when the evidence it affects is tangential rather than central and when the presentation threatens to obscure more than it can clarify. These are matters for the trial court to balance and juggle.¹²⁰ The general endorsement in *Branion* and *Statistics*

113. See, e.g., *PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE: THE USES AND LIMITS OF BAYESIANISM* (P. Tillers & E. Green eds. 1988).

114. Kaye, *The Laws of Probability and the Law of the Land*, 47 U. CHI. L. REV. 34 (1979).

115. Ellman & Kaye, *Probabilities and Proof: Can HLA and Blood Group Testing Prove Paternity?*, 54 N.Y.U. L. REV. 1131 (1979); see generally Kaye, *supra* note 114.

116. See authorities cited *supra* note 34.

117. *Branion v. Gramly*, 855 F.2d 1256, 1264 (7th Cir. 1988).

118. In *Branion*, the jury could understand the testimony about the driving and bruising times and appreciate, at least roughly, its implications without a mathematical analysis. This is not to deny that a suitable formal analysis could not aid the trier of fact, but it does mean that a probabilistic or decision theoretic analysis is not essential.

119. Bar-Hillel, *Probabilistic Analysis in Legal Factfinding*, 56 ACTA PSYCHOLOGICA 267, 282 (1984). As in *Branion*, "[j]udges could also check the sensitivity or robustness of the models. Such an analysis might show . . . that within a broad range it does not really matter precisely what the [value of a particular variable] really is." *Id.* at 283. Bar-Hillel advocates this "soft" role for decision theoretic modeling of the inferences from all the evidence in a case, which is a formidable task, *id.*, and one neither presented nor discussed in *Branion*.

120. A body of pertinent empirical research that promises to be of some assistance in this process is emerging. See Kaye & Koehler, *Can Jurors Understand Probabilistic Evidence?*, 154 J.

for Lawyers of "properly employed" statistical proof may incline some courts to be more open to probability calculations, but these remarks are dicta that dictate no concrete outcomes at the trial level.

I have treated *Branion v. Gramly* in detail not merely because it is an "interesting" opinion likely to find its way into the casebooks, but because it vividly depicts the danger of a little knowledge. The attempt to use probability theory in *Branion* was heroic. Like many acts of heroism, it also was hasty. Although there were some measurements and estimates of quantities bearing on guilt or innocence, the empirical data were so sketchy that the computations inevitably were more creative than convincing. Furthermore, counsel committed a few unforced errors. To this extent, the case brings to mind the aphorism that "[w]hen a lawyer acts as his own statistician, the performance is about what you would expect from a statistician who acts as his own lawyer."¹²¹

Courts as well as attorneys would do well to heed this admonition "before they start fooling with algebra."¹²² Despite a lapse or two in its own calculations, the *Branion* court successfully flagged the major problems in the sort of discourse that Auguste Comte once denigrated as "ponderous algebraic verbiage,"¹²³ but, in some respects, the algebra was more plausible than the court allowed, and in others, less so.¹²⁴ As Finkelstein and Levin indicate, other courts, also acting as their own statisticians, have fared less well.¹²⁵ Thus, a national panel

ROYAL STATISTICAL A. 75 (1991); Thompson, *Are Jurors Competent to Evaluate Statistical Evidence?*, LAW & CONTEMP. PROBS., Autumn 1989, at 9.

121. Carter, Flora, Van Bowen & Myers, *Statistics and the Virginia Blood Test Statute*, 56 VA. L. REV. 349, 357 (1970) (quoting Deming, *On the Contributions of Standards of Sampling to Legal Evidence and Accounting*, 19 CURRENT BUS. STUD. 14, 15 (1954)).

122. *Branion v. Gramly*, 855 F.2d 1256, 1264 (7th Cir. 1988). Moreover, most legal applications call for the skills of an applied statistician whose training extends beyond the mathematical theory of probability and statistics and encompasses some experience in working with data.

123. 4 A. COMPTE, COURS DE PHILOSOPHIE POSITIVE 366-68 (3d ed. 1809), reprinted in *PHYSIQUE SOCIALE: COURS DE PHILOSOPHIE POSITIVE, LEÇONS 46 À 60*, at 168-69 (Jean-Paul Enthoven ed. 1975).

124. Furthermore, one wonders whether the judges subscribing to the opinion in *Branion* could have joined some parts of that opinion on anything other than faith in its author or perhaps their law clerks. Justice Stevens was sensitive to this problem when he candidly stated in his dissenting opinion in *Hazelwood School District v. United States*, 433 U.S. 299, 318 n.5 (1977) (dissenting opinion), that "one of my law clerks advised me that . . . there is only about a 5% likelihood that a disparity this large would be produced by random selection from the labor pool." When passing on more traditional case or policy analysis, judges can verify the quality of the work for themselves. With most mathematical analyses, many judges would have to hedge their opinions with revealing qualifications like Justice Stevens' conditional conclusion, "[i]f [my clerk's] calculation is correct . . ." 433 U.S. 299, 318 n.5.

125. See *supra* note 25. For further illustrations, see, e.g., Fairley & Sugrue, *A Case of Unexamined Assumptions: The Use and Misuse of the Statistical Analysis of Castaneda/Hazelwood in Discrimination Litigation*, 24 B.C. L. REV. 925 (1983); Kaye, *supra* note 20; Kaye, *Statistical Evidence of Discrimination in Jury Selection*, in *STATISTICAL METHODS IN DISCRIMINATION LITIGATION*, *supra* note 7, at 13. In *Rowan v. Owens*, 752 F.2d 1186 (7th Cir. 1984), Judge Posner multiplied a few probabilities in *Collins*-like style to help demonstrate that evidence was sufficient to support a conviction. He argued only that his admittedly "arbitrary" numbers

recently warned of "possible dangers" when, as in *Branion*, "the judge is familiar with a statistical method the parties have not applied to data in evidence," and he does "his own ad hoc analysis."¹²⁶

Probability theory remains an arcane and dangerous tool for lawyers and courts. As *Statistics for Lawyers* argues, however, it is also a powerful and valuable instrument. The challenge for the legal system is to develop the knowledge and rules that will tend to produce technically sound and conceptually appropriate mathematical demonstrations, at trial or on appeal.

Statistics for Lawyers invests its resources, wisely, in the infrastructure of building knowledge rather than the analyses of doctrine and the construction of rules. The law student, attorney, or judge who appreciates the content of this textbook will be in a position to avoid or counter the kinds of errors made in *Branion* and to make more effective use of a broad spectrum of statistical thinking and resources. As Finkelstein and Levin explain, "[a] knowledgeable lawyer may not dispatch questions of legal policy with statistics, but by knowing more of the subject may hope to contribute to the store of rational and civilized discourse by which insights are gained and new accommodations reached. That . . . is the larger purpose of this book" (p. ix). It is a worthy purpose for a worthwhile book.

merely "bring out the point that it is wrong to view items of evidence in isolation when they point in the same direction." 752 F.2d at 1188. To this extent, his *sua sponte* use of probability theory is not troublesome.

126. STATISTICAL ASSESSMENTS, *supra* note 23, at 176. The panel recommended that in general, judges should not conduct analytical statistical studies on their own. If a court is not satisfied with the statistical evidence before it, alternative means should be used to clarify matters, such as a request for additional submissions from the parties or even, in exceptional circumstances, a reopening of the case to received additional evidence.

STATISTICAL ASSESSMENTS, *supra* note 23, at 176. Heeding this advice might have spared the *Branion* court the embarrassment of the arithmetical errors noted above, but would not have affected the outcome on appeal.